Applied Energy 233–234 (2019) 254–270

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

The feasibility and importance of considering climate change impacts in building retrofit analysis

Pengyuan Shen[a,*], William Braham[b], Yunkyu Yi[c]

[a] School of Architecture, Harbin Institute of Technology, Shenzhen 518055, China
[b] Department of Architecture, University of Pennsylvania, PA 19104, United States
[c] School of Architecture, University of Illinois at Urbana–Champaign, 61821, United States

HIGHLIGHTS

- Using 4% samples from the combinatorial space is sufficient for getting the most influential ECMs.
- Training on constructed extreme weather data enhances the predictive power of data-driven model.
- The optimal retrofit strategy of selecting the best ECM combinations is subject to climate change.
- Using the JMIM is able to find the most influential ECMs given a certain objective function.

ARTICLE INFO

Keywords:
Building retrofit
Climate change
Random forest
Feature selection
EnergyPlus
Energy use

ABSTRACT

Current building energy use is projected to increase by 1.7% annually until 2025, and the great potential for energy reduction in existing buildings has created opportunities in building energy retrofit projects. In this research, a framework and method are proposed to evaluate the impacts of different retrofit options to existing building under climate change. A Python retrofit tool is developed to perform parametric study by running EnergyPlus under different retrofit scenarios for existing buildings. With the help of Latin-hypercube sampling (LHS) method and a joint mutual information maximization (JMIM)-based feature selection method, the energy conservation measure (ECM) that may have the most potential in reducing the energy use or the lifecycle net present value (NPV) of a target existing building can be selected. A validated data-driven model is used to predict the building's future hourly energy use based on EnergyPlus simulation results generated by future extreme year weather data. It is demonstrated that global climate change will alter the optimal solution of future ECM combination and its influence varies from building to building, location to location. The optimal retrofit strategy of selecting the best ECM combinations under current climate condition will be subject to change in the future climate condition.

1. Background and introduction

In the United States, buildings consume 39% of the energy and 72% of the electricity, and they generate 38% of the carbon dioxide, 49% of the sulfur dioxide and 25% of the nitrogen oxides found in the air [1]. Contemporary buildings cannot function without the energy provided by fossil fuels, while pollutants from fossils fuels pose risks to the local and global environment with the growing needs of the development of human civilization. For decades, research attention has been paid to the energy efficiency of building, as existing buildings play an important role in building sector's energy consumption and carbon footprint. The method to find the best energy conservation measures (ECM) combinations for existing building retrofit has been studied and debated over years.

Regarding building retrofit, one of the focuses is to evaluate the building energy use (BEU) and the economic fitness with different combinations of ECM. This has been discussed in many studies under the current climatic conditions [2–9]. In recent studies, a commercial building energy retrofit analysis toolkit called Commercial Building Energy Saver (CBES) is proposed, which calculates the energy use of a building, identifies and evaluates retrofit measures in terms of energy savings, energy cost savings and payback based on the building energy simulation engine EnergyPlus and its user interface OpenStudio [10]. CBES is also used in city scale application of studying how much energy

* Corresponding author.
E-mail address: Pengyuan_pub@163.com (P. Shen).

https://doi.org/10.1016/j.apenergy.2018.10.041
Received 30 June 2018; Received in revised form 24 September 2018; Accepted 9 October 2018
Available online 18 October 2018
0306-2619/ © 2018 Elsevier Ltd. All rights reserved.

Nomenclature

ANN	artificial neural network
AFUE	annual fuel utilization efficiency
ASHRAE	American Society of Heating, Refrigerating, and Air-Conditioning Engineers
BES	building energy simulation
BEU	building energy use
CBES	Commercial Building Energy Saver
COP	coefficient of performance
CV	coefficient of variation
DOE	Department of Energy
ECM	energy conservation measure
IECC	International Energy Conservation Code
IPCC	Intergovernmental Panel on Climate Change
LCA	lifecycle cost analysis
LHS	Latin hypercube sampling
GCC	global climate change
GCM	global climate model
GHGs	greenhouse gases
GSHP	ground source heat pump
JMI	joint mutual information
JMIM	joint mutual information maximization
NPV	net present value
RCPs	representative concentration pathways
RE	renewable energy
RF	random forest
RMSE	root mean squared error
SCOP	seasonal coefficient of performance
SHGC	solar heat gain coefficent
SVM	support vector machine
TMY	typical meteorological year

can be saved for six city districts in northeast San Francisco by implementing a combination of five individual ECMs [11], and the impacts of modeling these commercial buildings in different ways, including different zoning methods and floor multipliers also have been studied [12]. However, recent efforts are paid to studying building retrofit problems under current climate condition. There still remains work to be done in finding a computationally efficient way of assessing the different energy performance provided by different ECM combination in future climatic condition. In this chapter, the background of the research subject, discussion of the state of the art, and the challenges will be discussed.

1.1. Impact of global climate change on building energy use

The level of greenhouse gases (GHGs) under different scenarios projected by the Intergovernmental Panel on Climate Change (IPCC) demonstrates a dramatic increase of GHGs in the future [13]. Many studies have shown that global climate change (GCC) brings great influences to BEU around the world. In Australia, Wang et al. evaluated the heating and cooling energy requirements and the corresponding carbon emissions of residential houses under different future climatic conditions [14], and found that the carbon emission of a 5-star house was projected to have an average increase of 30% in Darwin, 15% in Alice Springs and 19% in Sydney. Hassan Radhi assessed the potential impact of GCC on residential buildings in the United Arab Emirates with regard to CO_2 emissions [15]. In this research, the design of building envelope and fenestrations in the future is given highlight in combating the increase of building energy consumption as well as CO_2 emission. Shen stated that in four representative cities in the United States, the annual energy use is expected to change from -1.64% to 14.07% for residential buildings and from -3.27% to -0.12% for office buildings under A2 scenario (a carbon emission scenarios defined by IPCC) in different regions during 2040 to 2069 [16]. Two recent studies also reveal that GCC will affect the efficiency of building onsite renewable systems: one claims that GCC will bring down the energy efficiency of the GSHP system in residential applications in the United States using TRNSYS and eQuest modeling technique as the warmer ground in the future will result in an average rise of about 2–3 °C in the inlet and outlet water temperatures of ground source heat pump (GSHP) during the cooling season [17], while the other research indicates that for all the existing net zero energy buildings located in the ten climate zone in the United States, the proportion of decreased annual photovoltaic output is 5 out of 20, while that of wind turbines is 12 out of 20, indicating the comparatively lower stability of the renewable energy system prioritized to wind turbines than to photovoltaic prioritized renewable energy (RE) system [18]. The impacts of GCC on energy use in different types of buildings in the future are expected and analyzed in various research. Given the fact that the retrofitted building will last for a certain period of time during post-retrofit stage, it is essential that during the decision-making process, the situation of the possible change in building energy performance introduced by GCC as well as the retrofit measures should be studied. Therefore, it is necessary that GCC's implication to building performance should be studied since building energy retrofit will be made for the "future" building instead of buildings in current situation.

1.2. Challenges for building retrofit analysis under GCC

The building retrofit process can be complicated, involving the efforts of different parties including property owners, building clients, architects, and engineers. One of the most important issues facing building owners and clients is finding the best or the "bests" ECM combination for the existing building. This is a problem that the building energy research community has faced with for over thirty years. Asked from different perspectives, this question also means differently by saying the "best" or the "bests" because various emphasis is often put on building renovation: it could be the economic benefits brought by the renovation, or the enhancement of the building indoor environment and quality, or the reduction of BEU and GHGs. Thus, the optimization function is generally multi-objective and multi-constrained [2–4,19–21]. In this paper, the most imminent problem — how to analyze the retrofit project in a computationally affordable manner in future climate will be addressed when the decision-making factors of lifecycle economic analysis and energy performance are taken into account.

In recent years, the rapid development of computer technology and building energy simulation (BES) software has become a powerful tool in answering the question of how to integrate optimization approach with simulation results, which does not only provide a way in understanding BEU of different ECM combinations but also its environmental impacts and building wellness. Simulation becomes a powerful tool to achieve the goal — that is, to critically assess, refute and further propose new direction of future high performance building design. It also makes it possible to study the performance of independent subsystem and their interactive behavior between them, notably by using object-oriented simulation tools. Existing BES tools like EnergyPlus, eQuest, TRNSYS, ESP-r, etc. [22–24], can model and analyze building energy performance in terms of thermal dynamic features, day lighting, and human behavioral factors. Thus, various retrofit measures can be applied virtually to buildings and simulated results will inform users about the decision-making process in selecting ECMs. In general, computer-based simulation tools allow us to test different combinations of ECMs for a particular building in a relatively efficient and cost-effective way given a calibrated and accurate baseline model. However,

the use of BES for optimizing ECM building combinations with parametric case studies presents several challenges:

(a) First, the use of BES to evaluate building performance for the huge combinatorial problem when dealing with multiple ECMs brings an overwhelming computational cost. In a retrofit problem, each ECM and its parameter are considered as a vector of design variable. Due to the non-linear nature of optimizing building retrofit problem, it is hard to construct an objective function with derivative using method of gradient descent or ascent [25–27]. In order to find the interactions between different variables and their parameters, a huge combinatorial problem will be formed. The introduction of an additional ECM into a retrofit project will result in an exponential increase in simulation cases. Thus, using current optimization techniques to find solutions to the problem would be infeasible, even with parallel computing technology. The application of gradient-free and non-derivative optimization algorithms such as pattern search and genetic algorithm can be useful in solving the combinatorial optimization problem, but without a computationally efficient method for evaluating future year's building energy use, it is difficult to obtain optimal ECM combinations in the future.
(b) The huge computational cost in generating a training database for the data-driven model. Data-driven method can be used to significantly reduce computational cost in solving the building retrofit problems because it can be used to predict BEU of each ECM combination applied to the target building with potentially high reliability if data samples are sufficient and the model is properly trained. In a research that uses data-driven method to learn how different ECMs and their parameters will change the BEU, a huge database shall be established and used as the training database for the retrofit problem [28]. Nevertheless, if EnergyPlus is used to simulate the BEU of different combinations of ECMs under current climatic conditions, the enormous calculation cost is expected as mentioned in point (a) and also mentioned by Rysanek and Choudhary [8], not to mention running each case for the next 30 years with hourly weather data from each year generated by the future climate scenario. In fact, without the help of data driven model, it is rather impossible to perform the optimal ECM combination searching in this study. However, the problem of how to judiciously reduce the size of the training database and save computing power in database generation as well as in the training and forecasting process needs to be resolved.

1.3. Purpose of this study

Most research on existing building retrofit evaluation and analysis does not take into account the impacts of GCC. Meanwhile, the research on GCC's impact is generally focused on the energy performance of existing buildings. Few work has been done on the ongoing energy performance of existing buildings during the post-retrofit stage and the long-term performance of ECMs under climate change conditions. Recent research by Chow et al. studied the effectiveness of retrofitting existing public buildings in the face of GCC in the "hot summer and cold winter" climate region in China [29]. The study focuses mainly on existing buildings in Zhejiang Province. It uses the HadCM3 global climate model results to estimate the future building energy use by running the new weather file in DOE2 simulation engine, but only two retrofit measures are considered in this research. Another research in 2016 by Swedish and Swiss researchers showed a research closely related to building energy retrofit measures and theirs robustness to future climate change in terms of mainly heating energy consumption using a simplified model [30]. Limitations still exist in related studies, including very few ECM options, limitation of regional application, large simulation time step, and lack of an inclusive methodology that can be applied to any building type in various climate zones.

As building energy use is influenced during the retrofit lifetime, the forthcoming economic analysis is also subject to change when future climate conditions are considered in the retrofit. Challenges mentioned in Section 1.2 are confronted in this study and the proposed approach provides one of the resolution possibilities. The impacts of GCC on the ranked optimal ECM combinations regarding lifecycle net present value (NPV) will be assessed. This research is a first step towards future research on multi-objective decision making if economic or environmental factors for building retrofit are considered. The decision making of selecting ECMs for a building retrofit always concerns the cost-effectiveness of ECMs, this paper attempts to find if the best cost-effective scenarios would be affected by GCC. If so, then in a multi-objective optimization of building retrofit problem, the optimal solutions will also be affected, to a great extent, by BEU as well as the saving in the utility cost brought by the saving in BEU.

2. Approach

2.1. A Python based EnergyPlus parametric study tool

EnergyPlus is a widely used BES model in both academic and commercial studies, which is developed by the Department of Energy (DOE). Its precedent versions are BLAST and DOE-2 and it inherited the features and strengths of both programs. EnergyPlus is able to model the whole building energy performance and has undergone numerous reliability tests [31]. One of the characters of EnergyPlus is that it uses ASCII text based weather data, building input, and simulation output files and is able to handle design and engineering alterations by modifying the textual information in the input file. This important feature allows user to process inputs and transform the user inputs into text based information for EnergyPlus to override the baseline building model. In this way, different retrofit scenarios of implementing various ECM combination for the existing building can be realized in EnergyPlus model. The text based outputs generated by EnergyPlus can be collected and analyzed later. In this study, Python is used as the packaging tool for feeding EnergyPlus with retrofit inputs by user defined retrofit options. Python is good for text file processing, and has strong support and capability in data processing. It is chosen to program the main console and tool that helps ECM information injection and EnergyPlus output handling. All the retrofit-related IDF objects in EnergyPlus will be ready by the program and transformed into Python objects. Then alterations can be made to the Python objects and then have these Python objects transformed back to texted-based inputs for IDF file.

Building ECMs mainly fall into three major categories: passive, active, and renewables as shown in Fig. 1: (1) Passive retrofit is mainly about reducing the heating and cooling load of a building, including building envelope insulation, fenestration system retrofit, and natural ventilation. Passive retrofit ECMs aim to curb the heating and cooling load of the building when the building is supposed to maintain at a certain thermally comfortable condition for occupants; (2) Active retrofit refers to system efficiency, including HVAC system efficiency (heat source, pump, fan, valves, and etc.), lighting system efficiency, building control system. Active retrofit ECMs are applied to these building systems that are used to meet the energy demand caused by load as described in point 1 with decent efficiency; (3) Renewables refers primarily to the onsite renewable energy system, which provides additional or major portion of the required energy for the building in place of traditional fossil fuel energy source in order to reduce GHGs emission and total energy consumption. Renewables energies have attracted a great deal of interest in recent years, with the onsite renewable energy systems providing cleaner and less carbon-intensive energy supply alternatives.

In this study, we included the following ten ECMs: window replacement, wall insulation, roof insulation, window shading, air tightness improvement, cooling supply air temperature, cooling system efficiency upgrade, heating system efficiency upgrade, lighting efficiency,

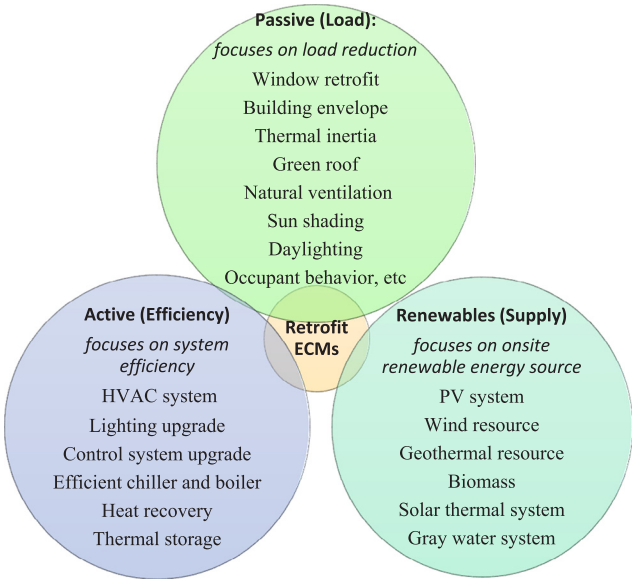

Fig. 1. Categories of building retrofit ECMs.

natural ventilation. The lifecycle of the retrofit is assumed to be from 2020 to 2060. The description of the ECMs are provided in Table 1, and the parameters and costs of all ECMs are listed from Tables 2–4 [2,11,20,32–34]:

2.2. Future local hourly weather data

In this research, the global climate model (GCM) — HadCM3, developed at the Hadley Center in the United Kingdom, is adopted to generate future weather file. HadCM3 (Hadley Center Coupled Model Version 3), like other GCMs, is a grid point model with large grid cells (2.5° in latitude and 3.75° in longitude over land areas, which gives 96 * 73 grid points on the scalar grid) [35]. The outputs of the HadCM3 model are monthly averages for each climatic variables of the chosen future period. This model has been used in some building energy research [36,37]. It is assumed that the building retrofit will last 40 years in this research. A recent run of HadCM3 model's outputs is used (IPCC Fifth Assessment Report (AR5)). For each GCM, the simulations were performed with prescribed CO_2 concentrations reaching 421 ppm (RCP2.6), 538 ppm (RCP4.5), 670 ppm (RCP6.0), and 936 ppm (RCP 8.5) by the year 2100 [38]. These scenarios are described in the IPCC Fifth Assessment Report (AR5) and are named representative concentration pathways (RCPs). In this paper, the RCP6.0 scenario is selected to predict future climate condition for the building retrofit as an attempt to find the potential impacts of climate change on building retrofit. In RCP6.0, emissions peak around 2080, then decline.

Due to the fact that GCM operates on a coarse global grid to model and predict the future GCC scenarios using a numerical solution to Navier-Stokes equations to save the computational resources, GCC outputs cannot be directly used as EnergyPlus weather data. The outputs need to be "downscaled" to the local geographic location with finer granularity and mapped to hourly step. Thanks to the times series based morphing downscaling method proposed by Belcher et al. [39], it is computationally efficient to obtain local hourly weather data in the future. The detailed morphing algorithm and its results in 10 climate zones in the United States are conducted and obtained in [18]. In this study, San Francisco and Philadelphia are chosen to be experimented with the proposed analysis approach. The reason for choosing these two cities for this study is because San Francisco has stable temperature spread over a year with lowest temperature at 10 °C in January and high temperature at around 27 °C in summer while the variance of outdoor dry bulb temperature in Philadelphia is high throughout the year (humid subtropical climate zone that has hot summer and cold winter) [40], making comparison for prioritizing different ECMs for buildings in the two climate zones interesting.

In the morphing method, there is a "baseline" where the "morphing" starts from and the baseline refers to the TMY3 weather data. The first step of this method is to calculate the mean value for each climate variable of each month m for the baseline scenario, the baseline climate value of x_0 for month m is defined to be:

$$\langle x_0 \rangle_m = \frac{1}{24 \times d_m} \sum_{monthm} x_0 \quad (1)$$

where d_m is the number of days in month m and the 24 comes from averaging the hourly measurements over the 24 h of each day.

The morphing method adopted here includes three operations, which can be described as: (1) a shift; (2) a linear stretch (scaling factor); (3) a shift and a stretch, the following equations demonstrates the three operations:

$$x = x_0 + \Delta x_m \quad (2)$$

$$x = \alpha_m x_0 \quad (3)$$

$$x = x_0 + \Delta x_m + \alpha_m (x_0 - \langle x_0 \rangle_m) \quad (4)$$

where x_0 is the existing hourly climate variable, Δx_m is the absolute change in monthly mean climate variable for month m (which is obtained from GCM outcome), α_m is the fractional change in monthly

Table 2
Window replacement properties and cost.

Window (SHGC, U-value (W/m^2 °C))	$/m^2	Type
(0.0, 0.0)	0.00	N/A
(0.80, 3.6)	47.0	Single glazing
(0.75, 2.8)	53.2	2bl glazing Without thermal break
(0.62, 1.6)	75.2	2bl glazing low-e window
(0.44, 1.6)	92.9	2bl glazing Window air-filled metallic frame
(0.288, 1.05)	79.2	SGSILVER
(0.585, 0.52)	98.1	SGCLIMATOP
(0.28, 0.33)	113.4	3050 SH 7/16 in. glass low-e
(0.63, 0.48)	131.7	3050 SH 7/16 in. glass
(0.25, 0.26)	183.0	3050 DH 3–7/16 insulated glass low-e krypton filled triple pane

Table 1
Description of ECMs.

Window replacement	change the window to another window type with different properties in thermal conductivity and solar heat gain coefficient
Wall insulation	adding insulation to the external walls of the building
Roof insulation	adding insulation to the external roofs of the building
Lighting efficiency upgrade	use energy efficient lighting equipment for all the zones in the building
Natural ventilation	open parts (%) of the windows in the building during swing seasons and during summer nights to flush redundant heat in the building
Window shading	use blinds shading for the windows when the building zone is overheated and needs cooling
Cooling supply air temperature	change the cooling supply air temperature
Heating efficiency improvement	replace existing heating system with high-efficiency heating source
Cooling efficiency improvement	replace existing cooling system with high-efficiency cooling source
Air tightness improvement	tighten up the building and improve the air change rate between indoor and outdoor environment

Table 3
ECM parameters and costs (w/cost).

Wall insulation (R value) (m^2 °C/W)	$/wall m^2	Roof insulation (m^2 °C/W)	$/roof m^2	Window shading	$/window m^2	Air infiltration (h^{-1})	$/$m^2$
N/A	0	N/A	0	N/A	0	N/A	0
1.25	11.4	1.52	12.5	1	28.7	0.3	25.5
1.61	12.5	1.97	16.4	2	37.2	0.5	20.2
1.97	13.5	2.42	20.1			0.7	14.4
2.33	14.6	2.87	22.9			0.9	9.3
2.69	15.7	3.32	26.8				
3.05	16.7	3.77	30.3				
3.41	18.5						
3.77	20.5						
Heating system upgrade (boiler)	$/kBTU-hr	Cooling system upgrade	$/ton	Lighting efficiency improvement	$/$m^2$		
Office: AFUE 95	42	Office: COP 6.27 (chiller)	477	N/A	0		
Residential: AFUE 95	84	Residential: SCOP 5.15 (roof top unit)	1586	30%	3		
				40%	1.9		

Table 4
ECM parameters and costs (w/o cost).

Cooling supply air temperature	natural Ventilated window open ratio
11	N/A
12	50%
13	100%
14	

mean climate variable for month m, and $\langle x_0 \rangle_m$ is the climate variable xo average over month m.

The adding of absolute change in monthly mean climate variable for certain month is called shift as described by Eq. (2). It indicates the mean value in baseline scenario experiences an absolute change, like the change in atmospheric pressure. A stretch is used when the change of certain variable is embodied as fractional change rather than absolute increment, like solar radiation which can never be a positive number at night (Eq. (3)). The combination of stretch and shift can be applied to those variables like dry-bulb temperature, where both fractional diurnal change and absolute increment take place, especially when taking into account the changes in both maximum and minimum daily temperatures (Eq. (4)). After applying the stretch and/or shift changes to specific climate variables, the future hourly weather data can be morphed on the basis of current TMY hourly weather data. The morphed climatic data in this research include dry-bulb temperature, relative humidity, and global horizontal irradiation. They are injected back to the epw weather file after morph is completed and then used by EnergyPlus.

2.3. The calculation of net present value (NPV)

NPV is used as the assessment criterion for the effectiveness of the ECMs in this research. In calculating NPV, the lifecycle cost analysis (LCA) method is used taking into account the future increase in cost and the discount rate.

- Calculation of retrofit investment:

$$I_0 = \sum_i A_{wall,i} * C_{wall-insu,i} + \sum_j A_{roof,j} * C_{roof-insu,j} + \sum_l A_{win,l} * C_{win,l} + \sum_n A_{win,n} * C_{shade,n} + A_{floor} * C_{infl} + C_{others} (\$) \quad (5)$$

where $A_{wall,i}$, $A_{roof,i}$, $A_{win,l}$, $A_{win,n}$ is the area (m^2) of building wall, roof, and windows where i^{th} wall insulation, j^{th} roof insulation, l^{th} window material, n^{th} shading material with certain cost ($), respectively. A_{floor} is the total floor area where infiltration level was improved, which has a cost of C_{infl} ($/$m^2$). Other costs include installing onsite renewable energy sources, upgrading building lightings, and etc.

It is assumed that all ECMs that have an initial investment will be maintained every five years, which will result in a periodic maintenance fee. Thus, the total retrofit investment can be obtained in the following equation:

$$I_{total} = I_0 + \begin{cases} \sum_k \frac{(1+\tau_m)^k * I_k}{(1+r)^k}, & k\%5 = 0 \\ 0, & otherwise \end{cases} (\$)$$

where τ_m (%) is the cost increase in maintenance fee of each year, r is the discount rate (%), and I_k is assumed to be proportional to the initial investment of k^{th} ECM.

The introduction of maintenance fees every five years is to intended to ensure that the applied ECM will operate as well as in the beginning of its life. Though the maintenance cost occurs every five years, the cost increase τ_m and discount rate r will always be taken account into account in the calculation on an annual basis. For the case study that will be discussed later, the maintenance cost is assumed to be 12% of the initial investment and is effective every five years.

- Calculation of energy saving in dollars:

The utility cost for each different type of energy source will be assumed to increase annually, and the total energy saving in dollars in the lifecycle would be:

$$S^u_{total} = \sum_k \frac{(1+\tau_u)^k * (C^u_{post,k} - C^u_{pre,k})}{(1+r)^k} (\$) \quad (6)$$

$$S_{total} = \sum_u S^u_{total} (\$) \quad (7)$$

where S^u_{total} ($) is the total saving for type of energy source u, τ_u (%) is the cost increase for type of energy source u, and $C^u_{post,k}$, $C^u_{pre,k}$ is the energy cost of type u during post-retrofit phase for the building with and without retrofit in year k. Finally, S_{total} can be obtained by aggregating the energy saving in dollars of each utility type. The assumption for the price of electricity and gas are $0.13/kWh and $3.54/thousand cubic feet ($0.125/$m^3$), respectively.

Then, the NPV can be obtained by:

$$NPV = S_{total} - I_{total} \quad (8)$$

2.4. Feature selection for ECM selection

For the ten proposed full set of ECMs in this research, if all of their combinations are considered in the building retrofit, then there will be numerous EnergyPlus simulation cases in total, not to mention the exponential growth of simulation cases when new ECMs are added in practice. In fact, for a specific existing building, there will be ECMs that are not so influential to the calculated NPV or BEU. A feature selection procedure making use of Joint Mutual Information Maximization (JMIM) is conducted before starting to spend great amount of computational resources to simulate all the possible combinations.

In this research, a feature selection algorithm based on information theory is used to search the most influential ECMs to an existing building in terms of the NPV or BEU. The JMIM uses joint mutual information and the "maximum of the minimum" approach that chooses the most relevant and influential features to the NPV or BEU. The algorithm is able to avoid overestimating the importance of the features compared with conventional joint mutual information (JMI) method [41]. To have an idea of the method, let's first start with the concept of entropy in information theory [42]:

$$H(X) = -\sum_{x\in X} p(x)\log p(x) \tag{9}$$

In Eq. (9), H(X) is the entropy of variable X, which concerns about the uncertainty level for its distribution and the average amount of distribution of information required to describe the random variable. For discrete variables, p(x) can be calculated by the proportion of the number of instants with value x in the total number of instants (N). Then we can calculate the conditional entropy H(X|Y) of two distributions as:

$$H(X|Y) = -\sum_{y\in Y} p(y) \sum_{x\in X} p(x|y)\log p(x|y) \tag{10}$$

Conditional entropy basically tells the amount of uncertainty left in X after Y is seen. Then mutual information I(X;Y) between X and Y is defined as:

$$I(X; Y) = H(X) - H(X|Y) \tag{11}$$

$$\sum_{x\in X} \sum_{y\in Y} p(xy)\log \frac{p(xy)}{p(x)p(y)} \tag{12}$$

The first term explains the uncertainty before Y is known, while the second term represents the uncertainty after Y is known. Mutual information can be thought of the amount of uncertainty in X that is removed by knowing Y. Then conditional mutual information can be defined as:

$$I(X; Y|Z) = H(X|Z) - H(X|Y, Z) = I(X; Y, Z) - I(X; Z) \tag{13}$$

where I(X;Y|Z) is the joint mutual information.

Lemma 1. *basically says that for a feature f_i, if the m-joint mutual information, which is $min_{f_S\in S}(I(f_i, f_s; y))$, is larger than that of all other features f_j, where f_i and $f_j \in F - S (i \neq j)$, then it is the most relevant feature to the target label y in the context of the subset of S. Here F is the full set of features, and let S be the subset of features that are selected already. JMIM employs joint mutual information and the 'maximum of the minimum' approach. It should choose the most relevant features according to* Lemma 1 [42]*, which is:*

$$f_{jMIM} = \arg max_{f_i\in F-S}(min_{f_S\in S}(I(f_i, f_s; y))) \tag{14}$$

where $I(f_i, f_s; y) = I(f_S; y) + I(f_i; y/f_S)$.

The algorithm starts with looking for the first feature that maximizes $I(y; f_S)$, then f_i will be removed from the full set F and enter the initially empty set S. Then the algorithm does not stop selecting the following features using the criterion described in Eq. (14) until the number of features in S reaches the desired number.

With the help of the JMIM algorithm, it is possible to find the most influential ECMs. Here "most influential" means most sensitive to the NPV or BEU of the retrofit project in the next 40 years. However, it is undoubted that the most influential ECMs can be identified by the ranking of the features when JMIM is run on the complete the database of all possible ECM combinations, but what is needed is to use fewer samples of simulation results to find out what these ECMs are. Therefore, we used Latin-hypercube sampling (LHS) technique to sample the combinatorial design space. Since LHS works only with continuous design space, the discrete retrofit design space is converted to continuous variable space, with each ECM parameter mapped to a uniform distribution. We then experimented with different sample sizes to find the minimum sample size that is sufficient to obtain a rough ranking of the feature importance. The results of the LHS sampling and random sampling are compared in this research when the method is applied to test case buildings, which will be presented in Section 3.2.

2.5. Data driven method for future year's energy use prediction

According to the challenge introduced in Section 1.2, the BES in future years of each ECM combination by any simulation-only procedure would be implausible due to the overwhelming computational cost. In this research, a data-driven method is used to predict the future years' BEU for each ECM combination after that the most significant ECMs are selected using the feature selection method proposed in Section 2.4.

The Random forest (RF) algorithm is used as a data mining method in this study. It is an ensemble learning method based on a non-parametric supervised learning method called a decision tree algorithm that uses a graph or tree model to learn and predict the schema of the best routes or rules. Compared with other popular data-driven algorithms (i.e. ANN: artificial neural network, SVM: support vector machine), RF has the advantages including less computational resources in model training, non-parametric traits that necessitates less parameter tuning (the number of ensemble tree used in this research is 110), and interpretability (ANN and SVM are not interpretable in terms of feature importance).

When applying the decision tree algorithm, the features of the independent variables can be either categorical or continuous. For a single decision tree, the simplest consistent explanation is the best, and such bias is called inductive bias. It is the set of assumptions that the learner uses to predict outputs given inputs that it has not encountered [43]. The basic algorithm for top-down induction of decision trees (ID3, C4.5 by Quinlan) is as follows [44]:

1. A, which is the "best" decision attribute for the next node.
2. Assign A as decision attribute for node.
3. For each value of A, create a new descendent of node.
4. Sort training examples to leaf nodes.
5. If training examples are perfectly classified, stop. Else, repeat over new leaf nodes.

In choosing the "best" attribute, the concept of entropy in Eq. (9) is also used in decision tree algorithm. Another important notion is Information Gain:

$$InfoGain(A, S) = H(S) - \sum_{t\in T} p(t)H(t) \tag{15}$$

where

- H(S) - Entropy of set S
- T - The subsets created from splitting set S by attribute A
- p(t) - The proportion of the number of elements in t to the number of elements in set S
- H(t) - Entropy of subset t

Furthermore, if the attribute has many values, information gain will

select it. Hence, in order to evade such node as date time like node, Gain Ratio should be used to evaluate the "best" attribute.

$$GainRatio(X, A) = \frac{InfoGain(X, A)}{SplitInformation(X, A)} \quad (16)$$

$$SplitInformation(X, A) = -\sum_{v \in value(A)} \frac{|X_v|}{|X|} \log_2 \frac{|X_v|}{|X|} \quad (17)$$

where X_v is a subset of X for which A has value v. Using this node division criterion when choosing the next division in the node, a tree-type model will be constructed. The more the tree develops in depth, the more complex the decision rules or routes and the fitter the model will be adapted to the training data. However, a single decision tree will be biased because it will overfit the training data, which means that if the hypothesis space has many dimensions (large number of attributes), meaningless regularity in the data that is irrelevant to the true, important, distinguishing features will be established. For example, a single decision tree will be sensitive to data outliers, making it fit the outliers and decreasing its predictive power, or if there is too little training data, even a reasonable hypothesis space will overfit. To settle this problem, RF is constructed by a set of decision trees in the training process and produces the rule that is the mean prediction of the individual trees. This corrects the overfitting behavior of a single decision tree. For RF, each tree in the ensemble is built from a sample drawn with replacement of the training set. When dividing a node during the construction of the tree, the chosen division is no longer the best split among all features. Instead, the split that is picked is the best split among a random subset of the features. Using RF instead of a single decision tree will effectively lower the variance of the model and handle the overfit problem.

In this study, RF is trained by the BEU database containing the hourly simulation results of selected influential ECM combinations based on future extreme year's hourly weather data constructed as described in Section 3.3. The features of the database include the weather variables including temperature, temperature change in the last hour, relative humidity, solar irradiation, wind speed, and building occupancy reported by EnergyPlus. Two different models are trained for electricity use and gas use of the building (if the building uses gas) for each building under certain future climate scenario (RCP6.0). Each retrofit plan will have its respective RF model for the prediction of future hourly BEU. The results and validity of the model will be discussed in Section 3.3.

3. Test and validation

3.1. Reference residential and office building

In this research, we use two reference EnergyPlus building models — one residential[1] and one office building[2] in the United States that are compliant with American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) 90.1's 2004 building code and International Energy Conservation Code (IECC) 2006 as the baseline case. San Francisco and Philadelphia are two climate zones where we apply and validate the proposed data-driven workflow and methods. It is assumed that the building retrofit lasts 40 years from the year of 2020 to 2060. Detailed descriptions of the building physical and thermal characteristics are shown in Table 5. The occupancy behavior, lighting and equipment intensity, and system schedule are shared for two types of buildings in San Francisco and Philadelphia respectively.

[1] https://www.energycodes.gov/development/residential/iecc_models.
[2] https://www.energycodes.gov/development/commercial/prototype_models.

Table 5
Major model parameter for the two buildings in San Francisco and Philadelphia.

	San Francisco		Philadelphia	
	Residential	Office	Residential	Office
Building Area (m^2)	335.0	4982.2	335.0	4982.2
Gross Wall Area (m^2)	235.1	1977.7	235.1	1977.7
Window Area (m^2)	33.2	652.6	33.2	652.6
Window U-value (W/m^2-K)	3.7	4.9	2.3	3.0
Window SHGC	0.4	0.4	0.4	0.4
Wall U-value (W/m^2-K)	0.5	0.8	0.5	0.8
Roof U-value (W/m^2-K)	2.5	0.4	2.5	0.4
Thermal Inertia (kJ/K m^2)	48.6	137.2	48.6	137.2
Air Infiltration rate (h^{-1})	1.5	1.5	1.5	1.5
Cooling Capacity (W)	7.18E+03	3.27E+05	1.45E+04	7.86E+05
Cooling COP	3.2	3.2	3.4	3.0
Heating Capacity (W)	1.23E+04	6.49E+05	2.26E+04	1.17E+06
Heating Efficiency	0.8	0.8	0.8	0.8
Lighting (W/m^2)	2.3	10.8	2.3	10.8
Service Water Power Input (W)	11137.8	29307.11	11137.8	29307.11
Service Water Heating Efficiency	0.81	0.81	0.80	0.81

3.2. Feature selection for training database size reduction

The proposed feature selection algorithm is applied to both types of building in two climate zones and we try to find the best sampling size with a decent chance to get the most influential ECMs for the target building without exhausting the whole combinatorial search space. To illustrate the validity of using the portion of the entire combinatorial space to obtain these ECMs, we choose a subset of 8 ECMs including window replacement, adding wall insulation, window shading, air tightness, heating system efficiency upgrade, lighting upgrade, cooling system efficiency upgrade, and natural ventilation, and experiment the validation of the feature selection method in the following steps:

(1) A "unit sampling size" concept is introduced here to describe the sample size by defining one "unit sampling size" as the number of all the ECM parameters considered in a retrofit, which means that, in this study, one unit of sampling size is 37 samples because there is a total number of 37 different individual parameters from all the ECMs. The total number of possible combinations is 48600 for the subset of 8 ECMs.
(2) LHS sampling is used to generate different number of sampling units ranging from one to forty for the continuous design space of ten dimensions, which has been mapped from the discrete space containing the ECM by uniform distribution between 0 and 1. We performed 200 times of feature selection for each number of sampling units and calculate the average accuracy of the feature importance rankings and compare it with the true feature importance rankings generated by the entire database.
(3) The same process in step 2 is conducted with random sampling to compare the feature selection accuracy of LHS.

The results of the feature selection test of using BEU as the objective function are illustrated in Fig. 2. The blue line and the red line in the figure corresponds to the LHS and uniform random sampling, respectively. For all the building types and climate zones, the use of LHS for feature selection has a great advantage over random sampling. Most importantly, it appears that using 30 sampling units of sample size can be sufficient to find the most influential ECMs can be thus obtained from the proposed ten ECMs in Table 5. Thirty sampling units of combinations represents approximately 3.7% of the total combinatorial search space. The proposed method of "LHS feature selection with 30 sampling units" makes it possible to focus on the most important ECMs that affects BEU, and reduce the computational cost greatly in the

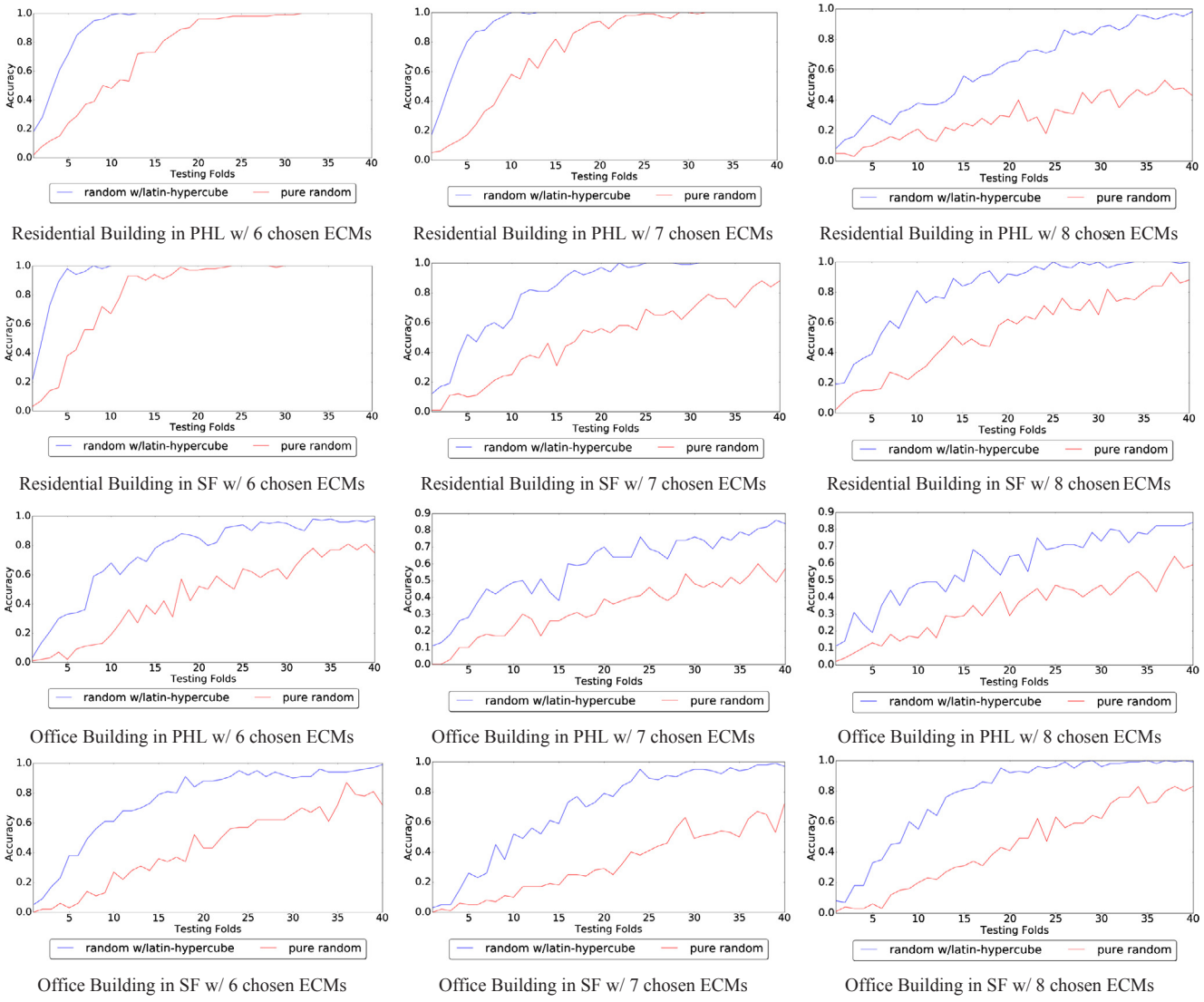
Accuracy
Testing Folds
random w/latin-hypercube
pure random
Residential Building in PHL w/ 6 chosen ECMs
Residential Building in PHL w/ 7 chosen ECMs
Residential Building in PHL w/ 8 chosen ECMs
Residential Building in SF w/ 6 chosen ECMs
Residential Building in SF w/ 7 chosen ECMs
Residential Building in SF w/ 8 chosen ECMs
Office Building in PHL w/ 6 chosen ECMs
Office Building in PHL w/ 7 chosen ECMs
Office Building in PHL w/ 8 chosen ECMs
Office Building in SF w/ 6 chosen ECMs
Office Building in SF w/ 7 chosen ECMs
Office Building in SF w/ 8 chosen ECMs

Fig. 2. Average feature selection accuracy with different sampling units by using LHS and random sampling (blue line: LHS, red line: uniform random sampling). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

simulation based parametric study.

To reduce the training database size and run simulations containing only the most important ECM combinations in EnergyPlus, the feature selection method is applied to the full set of ECMs for both building types and the 6 most influential ECMs are selected according to their feature importance against BEU. We also conducted the feature selection procedure using NPV as the objective to show the difference in the ECMs being recommended by the algorithm. The ranking of the selected ECMs for each building are shown in Tables 6 and 7. The objective function for feature selection here refers to the NPV and BEU calculated from the data-driven model predicted hourly BEU in future 40 years base on future "extreme year" hourly weather data. The reason, justification and method to generate the "extreme year" weather data will be introduced in Section 3.3.

3.3. Data-driven model training and validation

As described in Section 2.5, RF model is trained by the hourly BEU data generated by EnergyPlus. The features involved in the model training are: temperature, relative humidity (rh), solar irradiation, and the building occupancy schedule reported by EnergyPlus. Two models are generated to predict the hourly electricity use and gas use respectively. Then randomly sampled ECM combination is selected to validate the prediction results. We used TMY weather data to train the model, but it turned out that the prediction of the future BEU data is biased against the EnergyPlus simulation results in the place as shown in Fig. 3.

The predictive bias of the model is mainly due to the reason that the weather condition of the coming years are different from those of today. If one tries to predict the future BEU based on the model trained by

Table 6
Rankings of most influential ECMs for each building by BEU.

Residential in PHL	Office in PHL	Residential in SF	Office in SF
Air infiltration rate	Air infiltration rate	Air infiltration rate	Air infiltration rate
Wall Insulation (m^2K/W)	Cooling efficiency	Wall Insulation (m^2K/W)	Lighting efficiency
Window replacement	Window replacement	Heating efficiency	Window shading
Heating efficiency	Lighting efficiency	Window shading	Cooling supply air temperature (°C)
Window shading	Wall Insulation (m^2K/W)	Window replacement	Window replacement
Lighting efficiency	Window shading	Lighting efficiency	Cooling efficiency

Table 7
Rankings of most influential ECMs for each building by NPV.

Residential in PHL	Office in PHL	Residential in SF	Office in SF
Roof Insulation (m^2K/W)	Lighting upgrade	Roof Insulation (m^2K/W)	Lighting upgrade
Cooling efficiency upgrade	Air infiltration rate	Cooling efficiency upgrade	Air infiltration rate
Heating efficiency upgrade	Window replacement	Window replacement	Window replacement
Cooling supply air temperature (°C)	Window shading	Air infiltration rate	Window shading
Window replacement	Cooling supply air temperature (°C)	Heating efficiency upgrade	Natural ventilation
Lighting upgrade	Natural ventilation	Cooling supply air temperature (°C)	Roof Insulation (m^2K/W)

TMY weather data, the results will be biased. Thus, in order to improve the model prediction for future BEU, we try to construct a year of hourly weather data that contains the extreme weather conditions in the coming years and use that weather data to train the model. Then, EnergyPlus is used to simulate the hourly BEU based on the constructed weather file, and RF is then trained by the database with various ECM combinations. The end of this approach is to allow the regression model to better understand the variance in the future hourly weather conditions so that it can predict future hourly BEU more accurately.

Since temperature and building occupancy are the main driving forces of BEU for most types of building and the latter one is assumed not to change much in the future given that building occupancy schedule stays stable in future building operations, an "extreme year" weather data is constructed in terms of the extreme temperature in the future years. By means of concatenating half of each future weather data in the year with the most extreme winter (with the lowest temperature in 35 years) and the year of the most extreme summer (with the highest temperature in 35 years), the extreme year weather data is constructed. The construction process for Philadelphia is show in Fig. 4, where half year of the weather data in the year that has the hottest summer and the year that has the coldest winter has been concatenated to form a "pseudo year" hourly weather data. Due to that the constructed data represents a "pseudo year", the time frame may not be continuous.

The constructed extreme year hourly temperature in Philadelphia and San Francisco are shown in Fig. 5, where the blue line is TMY weather data and the orange line is the constructed extreme year weather data.

In addition, there are two parameters used in the RF model that determines the model performance of RF, which are the number of decision trees and maximum depth of each decision tree. The number of trees are set to be 100, and the maximum depth of each is 6 in this research since this combination gives a good cross validation results and in the meantime an acceptable computational cost.

Twenty cases are randomly selected among all the ECM combinations for each building to validate the prediction results by comparing the predictions with the EnergyPlus simulation results in the future years. The validation results of aggregated hourly BEU during the year of 2020 to 2060 are shown in Table 8. The root mean squared error (RMSE) and coefficient of variation (CV) of the model are used to evaluate the predictive power of the model. RMSE is frequently used to measure the difference between model predicted value and actually observed value, which can be taken as the residuals of the model, while CV is a description of the fitting of the model in terms of relative sizes of the residuals and outcome values. When CV gets lower, the smaller the residual is compared to the predicted values.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}} \tag{18}$$

$$CV = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}} \Big/ \bar{x} \tag{19}$$

where x_i and $\hat{x}_i$ is the true and forecasted value, $\bar{x}$ is the average of true values.

According to the results in Table 8, the model trained by future extreme weather data is performing well in predicting future aggregated hourly BEU. The highest CV for the annual BEU among all the cases in Table 8 is less than 3%, showing good reliability in annual BEU prediction.

Note: each row's results represent the prediction accuracy of the randomly chosen ECM combination in 2020 to 2060; it should also be noted that the twenty random ECM combinations differ for each building because they are randomly chosen for each building respectively.

Moreover, the results in Table 8 only reflect the model predictive

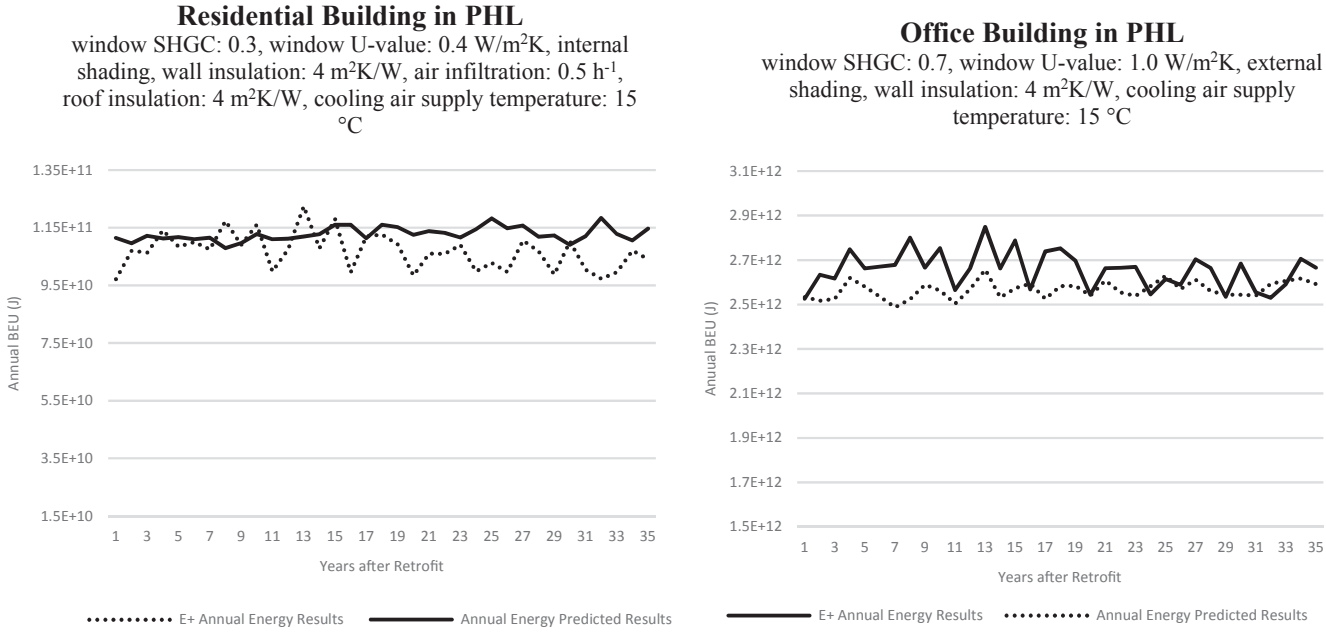
Residential Building in PHL
window SHGC: 0.3, window U-value: 0.4 W/m²K, internal shading, wall insulation: 4 m²K/W, air infiltration: 0.5 h⁻¹, roof insulation: 4 m²K/W, cooling air supply temperature: 15 °C
1.35E+11
1.15E+11
9.5E+10
7.5E+10
5.5E+10
3.5E+10
1.5E+10
Annual BEU (J)
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35
Years after Retrofit
E+ Annual Energy Results
Annual Energy Predicted Results
Office Building in PHL
window SHGC: 0.7, window U-value: 1.0 W/m²K, external shading, wall insulation: 4 m²K/W, cooling air supply temperature: 15 °C
3.1E+12
2.9E+12
2.7E+12
2.5E+12
2.3E+12
2.1E+12
1.9E+12
1.7E+12
1.5E+12
Anuual BEU (J)
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35
Years after Retrofit
E+ Annual Energy Results
Annual Energy Predicted Results

Fig. 3. Biased prediction results of the model for random ECM combination by using TMY weather data for training.

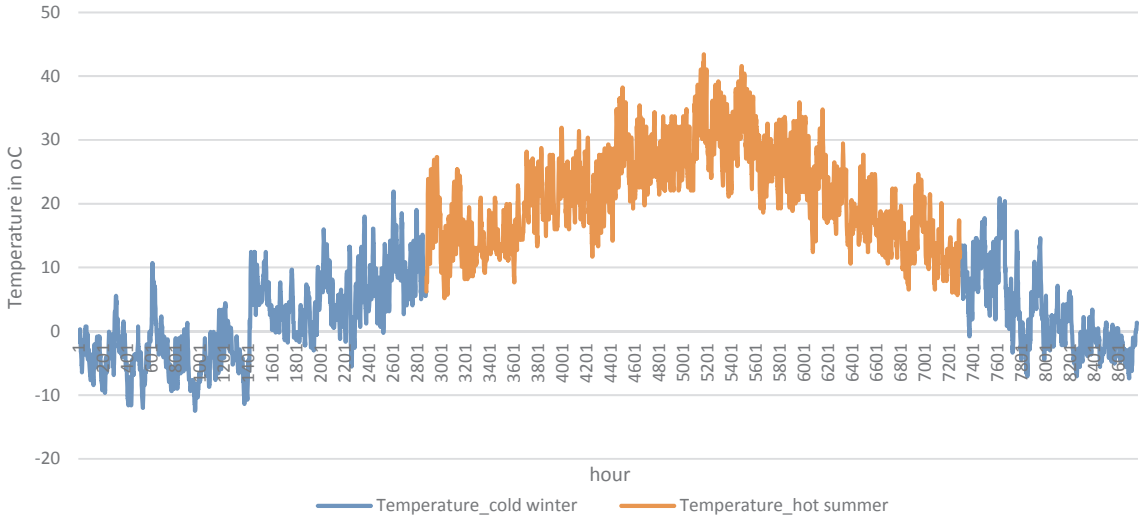

Fig. 4. The construction process of the extreme year hourly weather data (Philadelphia).

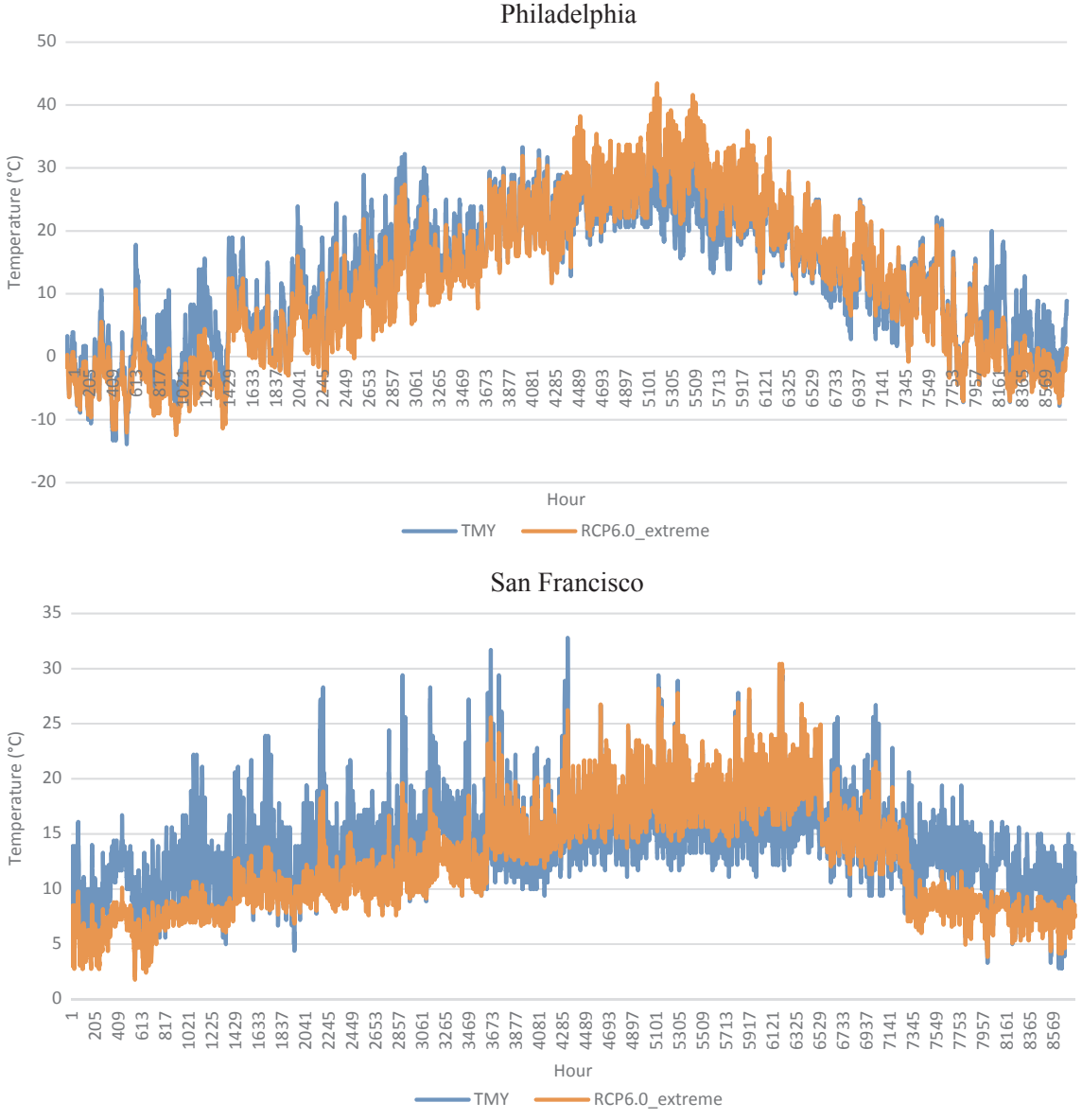

Fig. 5. The constructed hourly temperature profile for extreme weather year.

Table 8
Annual aggregated hourly BEU validation of twenty randomly chosen ECM combinations for each building in future years.

Residential building (PHL)		Office building (PHL)		Residential building (SF)		Office building (SF)	
BEU_RMSE (J)	BEU_CV (%)	BEU_RMSE (J)	BEU_CV (%)	BEU_RMSE (J)	BEU_CV (%)	BEU_RMSE (J)	BEU_CV (%)
1.18E+09	1.00%	7.45E+10	0.95%	1.06E+09	1.61%	5.57E+10	1.49%
1.40E+09	1.19%	6.40E+10	1.06%	8.73E+08	1.15%	9.07E+09	0.34%
1.25E+09	1.01%	4.57E+10	0.94%	5.07E+08	0.68%	4.47E+10	1.12%
1.23E+09	1.24%	8.66E+10	1.07%	1.04E+09	1.16%	1.35E+10	0.56%
1.14E+09	1.09%	6.87E+10	0.84%	1.01E+09	1.24%	3.42E+10	0.89%
1.32E+09	1.29%	8.06E+10	1.60%	1.86E+09	2.12%	5.71E+10	1.48%
1.52E+09	1.07%	7.34E+10	1.52%	9.89E+08	1.58%	5.37E+10	1.72%
1.76E+09	1.28%	6.75E+10	1.08%	1.16E+09	1.79%	3.56E+10	1.43%
1.46E+09	1.01%	6.59E+10	1.11%	5.83E+08	0.94%	8.30E+10	2.05%
1.40E+09	1.07%	5.35E+10	1.11%	4.26E+08	0.64%	6.70E+10	2.75%
1.23E+09	1.37%	7.84E+10	0.93%	5.42E+08	0.72%	8.34E+10	2.81%
1.95E+09	1.20%	6.52E+10	0.84%	1.01E+09	1.35%	8.00E+09	0.28%
1.48E+09	1.00%	8.31E+10	1.03%	1.04E+09	1.37%	6.25E+10	1.78%
1.88E+09	1.58%	7.88E+10	1.26%	1.41E+09	1.55%	4.75E+10	1.55%
1.39E+09	1.03%	5.29E+10	1.15%	1.40E+09	1.89%	5.30E+10	1.87%
1.42E+09	1.27%	8.31E+10	1.06%	1.08E+09	1.70%	5.89E+10	1.99%
1.72E+09	1.10%	5.10E+10	0.99%	7.80E+08	1.01%	4.01E+10	1.72%
1.88E+09	1.23%	4.34E+10	0.87%	6.01E+08	0.95%	7.23E+10	2.34%
1.18E+09	0.97%	9.08E+10	1.15%	8.12E+08	1.25%	6.25E+10	2.33%
1.35E+09	1.25%	6.17E+10	1.28%	9.49E+08	1.28%	6.65E+10	2.05%

power for aggregated hourly BEU. To show the performance of the RF model in predicting hourly BEU under future climate condition, the validation results of the hourly BEU in Table 9 are presented. As indicated in the table, the black box model is able to predict the hourly energy use of each different building type with most CV smaller than 8%. One of the reasons of why the RF model can do quite a good work in predicting hourly and annual energy use for the selected buildings in this research is that the four buildings are reference building and are supposed to be operated under good condition. The system degradation and the occurrence of system faults in practical operation have not been included in the training data and not considered by the data-driven model.

4. Results and discussions

By applying the proposed method to the two typical building types in Philadelphia and San Francisco, influential ECMs are selected from the ten ECMs by their feature importance using NPV and BEU (for comparison) as the objective for each building before the generation of the complete database by running the EnergyPlus simulation using the LHS feature selection with 30 sampling units, and the future hourly BEU for the year of 2020 to 2060 of each selected ECM combination is projected by RF model. Since the proposed method is validated in Sections 3.1 and 3.2, final results are generated and analyzed for each building.

4.1. ECM selections for each building

Six top NPV-sensitive ECMs are selected using the LHS feature selection with 30 sampling units for each building in Table 7. In addition, the six top BEU-sensitive ECMs are also presented in Table 6 to compare with the results in Table 7. The proposed JMIM feature selection method works well in identifying the most important ECMs for a particular building and can be applied in future related research. In order to better illustrate the results shown in Table 6, the end use break down of BEU for each building is plotted in Fig. 6. It should be noted that the

Table 9
Hourly BEU validation of twenty randomly chosen ECM combinations for each building in future years.

Residential building (PHL)		Office building (PHL)		Residential building (SF)		Office building (SF)	
BEU_RMSE (J)	BEU_CV (%)	BEU_RMSE (J)	BEU_CV (%)	BEU_RMSE (J)	BEU_CV (%)	BEU_RMSE (J)	BEU_CV (%)
1.28E+05	5.2%	8.03E+06	6.5%	1.11E+05	5.7%	7.71E+06	6.6%
2.62E+05	7.1%	1.14E+07	5.1%	1.47E+05	5.3%	1.17E+06	3.7%
1.39E+05	5.9%	4.77E+06	6.4%	8.22E+04	4.4%	7.11E+06	5.3%
2.17E+05	7.0%	9.05E+06	5.9%	1.40E+05	5.4%	1.39E+06	4.2%
1.09E+05	5.9%	8.38E+06	5.7%	9.22E+04	5.5%	3.35E+06	4.4%
1.71E+05	6.1%	1.54E+07	7.3%	1.80E+05	7.5%	6.47E+06	6.5%
2.24E+05	5.7%	7.99E+06	6.3%	1.74E+05	6.8%	5.82E+06	7.3%
1.79E+05	6.9%	8.73E+06	6.3%	1.09E+05	5.5%	3.73E+06	4.8%
2.01E+05	5.8%	9.00E+06	4.4%	9.25E+04	5.2%	1.04E+07	5.6%
1.55E+05	6.7%	6.84E+06	6.0%	5.72E+04	4.1%	1.22E+07	6.8%
2.12E+05	4.6%	1.08E+07	5.7%	9.72E+04	6.4%	8.94E+06	7.2%
2.22E+05	4.3%	9.78E+06	5.7%	1.88E+05	6.8%	8.59E+05	4.1%
1.50E+05	4.2%	7.65E+06	6.0%	1.21E+05	6.4%	8.71E+06	7.6%
2.18E+05	5.6%	7.18E+06	5.4%	1.64E+05	5.8%	8.75E+06	6.9%
1.61E+05	4.1%	5.87E+06	4.5%	1.79E+05	7.1%	7.86E+06	5.0%
1.41E+05	6.2%	1.02E+07	5.9%	1.84E+05	6.3%	1.13E+07	7.5%
1.62E+05	7.0%	4.84E+06	6.3%	1.07E+05	4.4%	7.01E+06	5.1%
2.34E+05	6.1%	6.83E+06	5.3%	6.41E+04	4.1%	9.82E+06	6.7%
1.66E+05	6.1%	1.31E+07	5.6%	8.63E+04	6.5%	6.62E+06	7.4%
1.92E+05	5.5%	1.01E+07	6.2%	1.62E+05	6.7%	6.71E+06	6.7%

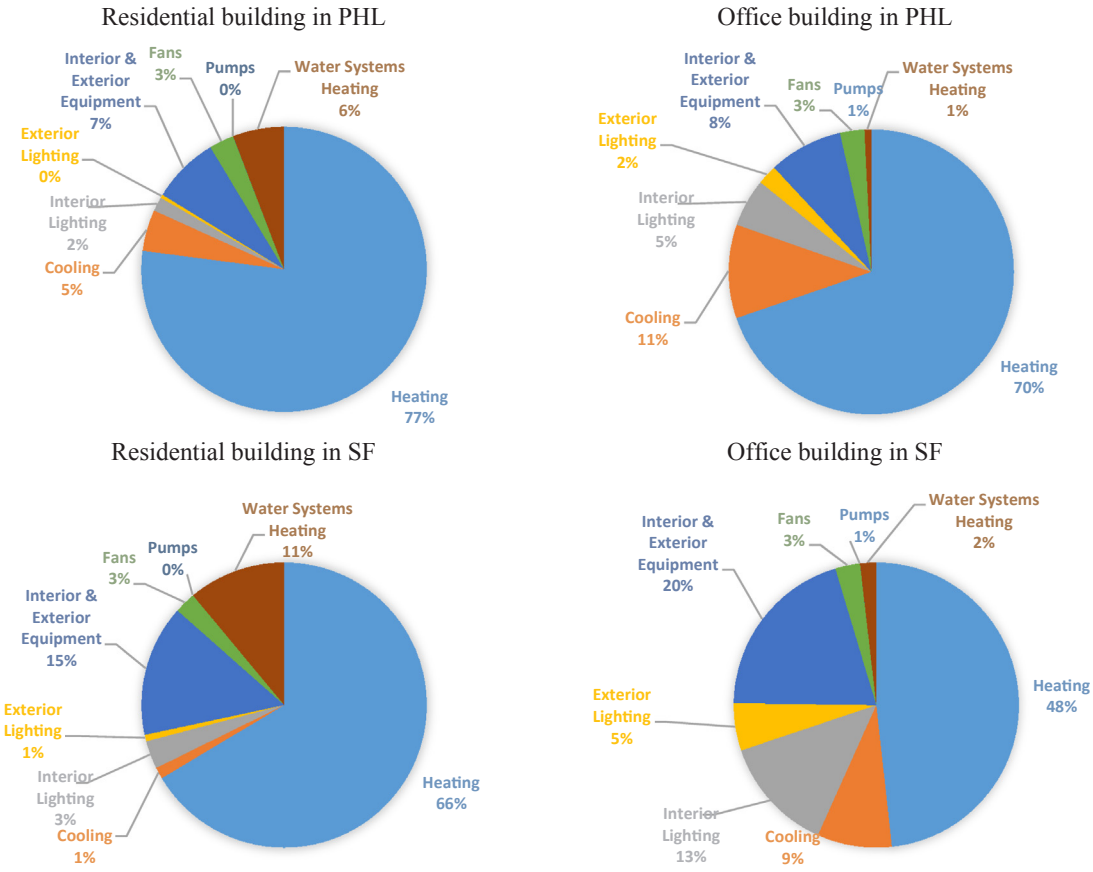

Fig. 6. End use breakdown of BEU for the four buildings.

reason why heating energy use is taking most of the percentages in total BEU is that the heating site energy source is gas for both types of buildings and gas has a lower site to source energy conversion factor compared to electricity. Usually, the site to source energy conversion factor of electricity is three times higher than gas in the United States [31]. As per the results, air infiltration is the most influential factor that determines the amount of BEU, regardless of building type, indicating that a different airtightness level can greatly influence BEU performance. For residential building, wall insulation and heating efficiency are important, which implies that the reduction of heating energy use in residential building is vital. It can also be told from Fig. 6 that heating energy use in residential buildings has a higher proportion in total BEU than office buildings, which is logical due to that BEU in residential building is more vulnerable to heat loss than heat gains and that office buildings are more influenced by higher thermal capacity, greater intensity of equipment use and occupant activity. This is also validated by the fact that the improvement of cooling efficiency is among the top 6 for the two office buildings in Table 6.

For both types of buildings, the improvement of lighting efficiency is more important in office buildings than in residential buildings because lighting energy use in office building has a higher proportion in the end use breakdown.

In this study, window shading is achieved by installing blinds inside or outside the window and both installations are considered and analyzed. Window shading is also among the “top 6” ECMs for all the four buildings. Window replacement is a factor that can never be underestimated in impacting BEU for both types of building, and the same thing happens for window shading because it is known that the heat transfer through the window is always decisive for the building heating and cooling energy use. Moreover, as indicated from the results in Table 6, window replacement is more important to the improvement of BEU for building in Philadelphia. This may be due to the fact that for such a colder climate as Philadelphia is located in, a better insulated glazing system is required than in San Francisco. In addition, the heating efficiency of residential building should receive more attention than cooling, while improving the cooling efficiency of office buildings should be emphasized.

When it comes to the six most influential ECMs in terms of NPV for the four buildings, the scenario changes. Since selecting the most influential ECMs according to NPV will be greatly subject to the cost effectiveness of each ECM to the building retrofit project during the lifecycle, the results will be highly sensitive to the ECM cost data. In this research, we collected some ECM cost data from related researches as shown in Tables 2 and 3 and conducted the analysis using the JMIM method. It is shown that for residential buildings, roof insulation is the most important ECM impacting the lifecycle cost of the retrofit. The ECMs that appear for both residential buildings include cooling and heating efficiency upgrade, window replacement, and cooling supply air temperature optimization.

For office buildings, lighting upgrade becomes the most important factor influencing the NPV of the retrofit. The reduction in lighting intensity will not only cut down the electricity consumption of the lighting system per se, but also potentially reduce the cooling need for the type of building such as office building, which has a large amount of internal heat gain. The economic return from electricity reduction due to the lighting upgrade could be substantial. Replacing the window and tighten up the building airtightness are also crucial to the NPV of the retrofit. More importantly, the inclusion of natural ventilation also indicates that the removing of the internal heat gain will bring financial benefits to the building in the future climate. The same inference also applies to the selection of window shading for office buildings. The other selected ECMs for the office buildings include cooling supply air temperature change and roof insulation.

4.2. Impact of climate change on hourly BEU

To show the impact of climate change on the hourly BEU, we randomly select an ECM combination for each building and compare their total energy demand in the year of 2020 and 2060. Moreover, for each building, the baseline scenario to which no ECM being applied is also included. The energy demand in a randomly selected summer and winter week for those scenarios are presented in Fig. 7:

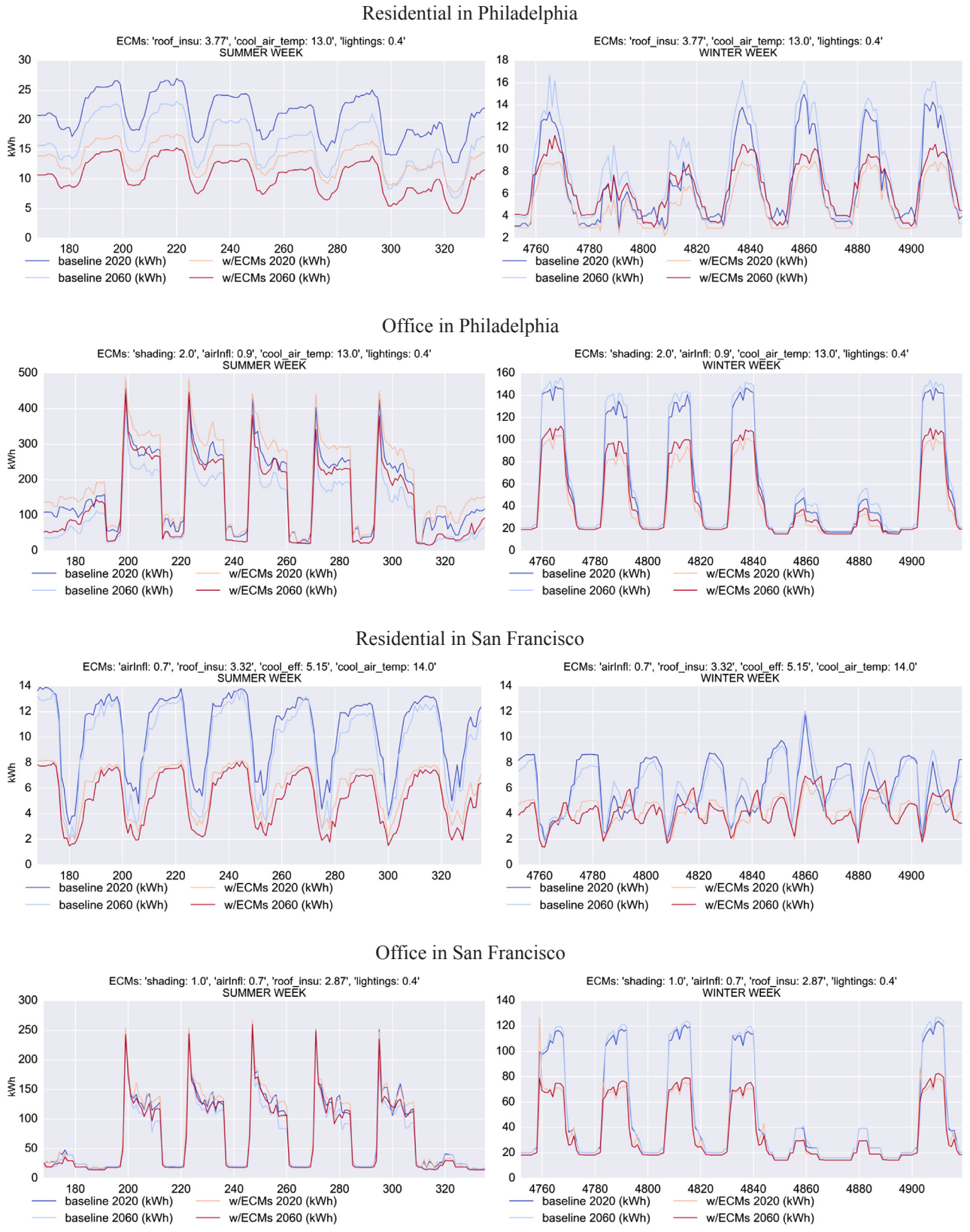
Residential in Philadelphia
ECMs: 'roof_insu: 3.77', 'cool_air_temp: 13.0', 'lightings: 0.4'
SUMMER WEEK
WINTER WEEK
kWh
baseline 2020 (kWh)
w/ECMs 2020 (kWh)
baseline 2060 (kWh)
w/ECMs 2060 (kWh)
Office in Philadelphia
ECMs: 'shading: 2.0', 'airInfl: 0.9', 'cool_air_temp: 13.0', 'lightings: 0.4'
Residential in San Francisco
ECMs: 'airInfl: 0.7', 'roof_insu: 3.32', 'cool_eff: 5.15', 'cool_air_temp: 14.0'
Office in San Francisco
ECMs: 'shading: 1.0', 'airInfl: 0.7', 'roof_insu: 2.87', 'lightings: 0.4'

Fig. 7. Impact of climate change on the hourly BEU in the four cities (randomly selected ECM combination).

The impact of climate change on the hourly building energy demand is visible in the four reference buildings. For the residential building in Philadelphia, the energy use demand during winter in the year of 2020 and the year of 2060 can be very different for both the baseline and retrofitted scenario. For the residential building in San Francisco, the impact of climate change on the hourly BEU is not too distinguished but still exists. The outdoor temperature in San Francisco is relatively cool, therefore the hourly BEU during the summer week does not have a regular pattern after the ECMs are applied to the building. It can be observed that the impact of climate change to the baseline and the retrofitted scenario is different.

For the two office building, the randomly chosen ECM combination includes the upgrading of lighting efficiency and window shading. It is illustrated in the figure that the energy use during summer has been greatly reduced while the energy use during winter has a small amount of increase. During the summer week, the impact of climate change on hourly BEU mainly occurs at peak load time and valley load time. The impact is stronger for the office building in Philadelphia during the winter than in San Francisco. In Philadelphia, the 2060 scenario will use less energy than in 2020 in winter and it will use more energy than

Table 10
Rankings of top 10 ECM combinations from year of 2020 to 2060 according to NPV.

Residential in PHL

Window	Lightings (%)	Cooling efficiency	Roof insulation (m^2K/W)	Cooling air temperature (°C)	Heating efficiency	NPV ($)	Electricity saving (kWh/m^2 yr)	Gas saving (kWh/m^2 yr)
(0.585, 0.52)	0.3	5.15	2.42	11	0.95	24102.1	−876.0	−5014.7
(0.585, 0.52)	0.3	5.15	2.87	11	0.95	24094.1	−888.6	−5121.0
(0.585, 0.52)	0.3	5.15	1.97	11	0.95	24064.0	−858.8	−4867.5
(0.28, 0.33)	0.4	5.15	2.42	11	0.95	23810.6	−843.0	−5032.8
(0.585, 0.52)	0.3	5.15	3.32	11	0.95	23805.3	−898.3	−5201.3
(0.585, 0.52)	0.4	5.15	2.87	11	0.95	23798.7	−855.4	−5138.9
(0.585, 0.52)	0.4	5.15	1.97	11	0.95	23771.8	−825.8	−4885.8
(0.28, 0.33)	0.3	5.15	1.52	11	0.95	23769.6	−833.4	−4650.4
(0.28, 0.33)	0.4	5.15	3.32	11	0.95	23505.7	−864.9	−5219.9
(0.585, 0.52)	0.3	5.15	3.77	11	0.95	23495.2	−905.7	−5264.7

Office in PHL

Window	Lightings (%)	Window shading	Cooling air temperature (°C)	Air infiltration rate (h^{-1})	Natural ventilation	NPV ($)	Electricity saving (kWh/m^2 yr)	Gas saving (kWh/m^2 yr)
(0.288, 1.05)	0.3	2	11	0.5	1	637651.7	−1126.4	478.7
(0.288, 1.05)	0.3	2	12	0.5	1	632214.9	−1117.9	478.7
(0.288, 1.05)	0.3	2	13	0.5	1	626219.0	−1108.6	478.7
(0.288, 1.05)	0.3	2	14	0.5	1	618073.7	−1096.0	478.7
(0.288, 1.05)	0.3	1	11	0.5	1	610988.3	−1071.5	475.8
(0.585, 0.52)	0.3	0	11	0.5	1	610941.1	−1026.7	472.2
(0.585, 0.52)	0.3	1	12	0.3	0.5	605081.8	−1062.3	475.8
(0.585, 0.52)	0.3	0	12	0.5	0.5	604671.8	−1017.0	472.2
(0.62, 1.6)	0.3	2	11	0.5	1	595558.6	−1128.9	−108.8
(0.62, 1.6)	0.3	2	12	0.5	1	590376.8	−1120.9	−108.8

Residential in SF

Window	Cooling efficiency	Roof insulation (m^2K/W)	Cooling air temperature (°C)	Air infiltration rate (h^{-1})	Heating efficiency	NPV ($)	Electricity saving (kWh/m^2 yr)	Gas saving (kWh/m^2 yr)
(0.585, 0.52)	5.15	1.97	11	0.9	0.95	17730.2	−367.1	−3687.7
(0.585, 0.52)	5.15	2.42	11	0.9	0.95	17697.2	−377.4	−3793.9
(0.585, 0.52)	5.15	2.87	11	0.9	0.95	17637.2	−384.9	−3871.7
(0.585, 0.52)	5.15	1.52	11	0.7	0.95	17545.0	−352.0	−3529.7
(0.585, 0.52)	5.15	3.32	11	0.9	0.95	17291.9	−390.4	−3927.7
(0.28, 0.33)	5.15	3.77	11	0.9	0.95	16938.1	−394.5	−3972.6
(0.288, 1.05)	5.15	1.97	12	0.7	0.95	16847.0	−340.2	−3648.1
(0.585, 0.52)	5.15	2.42	12	0.9	0.95	16826.3	−351.0	−3754.0
(0.28, 0.33)	5.15	2.87	12	0.5	0.95	16777.9	−359.0	−3830.1
(0.288, 1.05)	5.15	1.52	12	0.7	0.95	16618.9	−324.0	−3486.5

Office in SF

Window	Lightings (%)	Roof insulation (m^2K/W)	Window shading	Air infiltration rate (h^{-1})	Natural ventilation	NPV ($)	Electricity saving (kWh/m^2 yr)	Gas saving (kWh/m^2 yr)
(0.75, 2.8)	0.3	0	2	0.7	1	615091.9	−1083.7	−392.6
(0.75, 2.8)	0.3	1.52	0	0.7	1	611,308	−1017.7	−367.1
(0.62, 1.6)	0.3	0	1	0.9	1	603348.5	−1050.9	−378.8
(0.75, 2.8)	0.3	1.52	2	0.7	1	585419.9	−1085.8	−380.9
(0.75, 2.8)	0.3	1.52	0	0.7	0.5	581849.1	−1020.2	−355.6
(0.0, 0.0)	0.3	1.97	2	0.5	1	575,732	−1086.2	−379
(0.62, 1.6)	0.3	1.52	1	0.7	1	573863.3	−1053.3	−367.2
(0.0, 0.0)	0.3	1.97	0	0.7	0.5	572254.9	−1020.7	−353.2
(0.62, 1.6)	0.3	0	2	0.9	1	568017.1	−1083.7	−209.7
(0.75, 2.8)	0.3	2.42	2	0.7	1	566818.6	−1086.5	−377.8

Note: In this table, “0” means that the specific ECM is not applied to the building.

Table 11
Ranking changes of ECM combinations for four buildings compared with TMY scenario.

		San Francisco		Philadelphia	
ECM	Parameter	Residential	Office	Residential	Office
Window	N/A	28.2	22.5	−18.5	−41.9
	(0.25, 0.26)	43.7	7.7	23.5	36.7
	(0.28, 0.33)	32.4	21.4	18.7	27.2
	(0.63, 0.48)	32.3	12.3	18.2	29.1
	(0.585, 0.52)	−23.1	−10	15.2	20
	(0.288, 1.05)	14.2	19.3	4.1	5.1
	(0.44, 1.6)	−2.8	−0.8	−1.3	5.4
	(0.62, 1.6)	−2.6	5.4	−3.6	−3.4
	(0.75, 2.8)	−43.3	−23.9	−21.6	−32.4
	(0.8, 3.6)	−79	−54	−34.8	−45.9
Window shading	N/A	–	12.8	–	−19.7
	Internal Blind	–	4.5	–	3.1
	External Blind	–	−17.3	–	16.6
Air infiltration rate	N/A	−58.5	76.8	–	−4
	0.3	53.7	116.2	–	96.5
	0.5	6.3	−64.7	–	35.1
	0.7	36.4	44.1	–	−33.2
	0.9	−37.9	−172.3	–	−94.4
Roof insulation	N/A	−94.9	10.9	−135.6	–
	1.52	4.6	5	6.2	–
	1.97	11.4	1.7	15.1	–
	2.42	16	−0.9	22.4	–
	2.87	19.1	−3	27.2	–
	3.32	20.6	−5.5	30.9	–
	3.77	23.2	−8.2	33.8	–
Heating efficiency	N/A	−6.3	–	−16	–
	0.95	6.3	–	16	–
Cooling efficiency	N/A	1.6	–	−38.5	–
	5.15	−1.6	–	38.5	–
Cooling air temperature	11	−6.2	–	18.2	7.4
	12	−3.9	–	7.6	3.1
	13	1.8	–	−6.6	−2.1
	14	8.4	–	−19.2	−8.4
Natural ventilation	N/A	–	−43.2	–	−25.6
	0.5	–	24.5	–	11.2
	1	–	18.7	–	1.1
Lighting efficiency	N/A	–	−1	−38.7	0
	0.3	–	83.3	24	41.2
	0.4	–	−82.3	14.6	−41.2

Note: "–" means that the ECM is not chosen by JMIM feature selection or does not apply to the specific building type.

in 2020 in summer.

4.3. Rankings of ECM combinations in future years

As described in Section 3.3, eight RF models are trained in total, two for each building being trained by electricity use and gas use as predictands, whereas future extreme year weather data and ECM parameters as predictors. The hourly BEU for each combination of selected ECMs is projected from 2020 to 2060. Ranking by their NPV during the year of 2020 to 2060, the top 10 ECM combinations that have the most NPV in the future 40 years are presented in Table 10.

Based on the results in Table 10, to have the best NPV, the best cooling supply air temperature for the residential buildings in Philadelphia and San Francisco would be 11 °C. Since it is assumed to be a free ECM, its energy saving will all be turned into financial saving. The cooling system upgrade is also very important for the financial performance during the retrofit period. More importantly, compared with the office building, the heating efficiency upgrade is unanimously adopted for the residential buildings. As it is discussed in the last section, since residential buildings are more susceptible to outdoor environment and have less internal heat gain, the efficiency of the heating system should be paid emphasis to make more profit for the retrofit project even under the climate change scenario.

Another important ECM for the financial performance of the residential buildings is roof insulation. Roof insulations with different R-values all have a chance to appear in the top 10 ECM combinations because of the synergy of various combined thermal performance and financial cost. The difference in adopting different R-values may have very little difference in the final calculated NPV, for example, for the top three ECM combinations in Philadelphia. However, the roof insulation is definitely needed to improve the financial performance of the residential buildings in both cities.

For office buildings, the upgrading of the lightings is the most important measure to produce financial return for the retrofit project. The two office buildings in both cities all benefit from lighting upgrading in terms of NPV due to that the reduction of the lighting intensity not only decrease the electricity consumption per se but also reduces internal heat gain for the office building which saves cooling energy use during most time of the year under the situation of global warming. Window shading is also another important measure to reduce cooling load for office building when future mean outdoor temperature has a rising trend.

The other two ECMs for office buildings are the air infiltration rate and the natural ventilation. Natural ventilation will bring benefits to the economic efficiency of the retrofit in the future because it removes redundant heat during the night time or when the outdoor air condition is suitable for passive ventilation for the office building, and most of the time, it can be free. However, for air tightness, the office building in Philadelphia will need a comparatively lower infiltration rate to pull up the NPV compared with that in San Francisco mainly due to the colder winter in Philadelphia. A better airtightness will make the building lose less heat when the outdoor temperature is low because it prevents too much air exchange between indoor and outdoor environment. Nevertheless, it can be seen that the strictest airtightness level $0.3\,h^{-1}$ is not among the top in the list because too much airtightness will make the retrofit less lucrative during its 40 years lifecycle.

It is interesting to look at the choosing of the windows in the retrofit of the two office buildings. The office building in Philadelphia prefers windows with comparatively lower U-value and lower SHGC, while the building in San Francisco prefers windows with higher U-value and higher SHGC. This might be attributed to the fact that the climate condition in San Francisco is more amenable to that in Philadelphia. The office building in Philadelphia needs a more isolated building operational condition against the outdoor climate than San Francisco to save more energy and thus have more NPV during its lifecycle even with a higher investment cost for those chosen window types.

4.4. Ranking change of ECMs in the future

Analyzing future top ECM combinations is not sufficient to understand the impacts of GCC on the decision-making process for building retrofit. We also compared the ranking changes of each ECM combination in terms of NPV in future 40 years against using TMY as the weather input. We used EnergyPlus to run the selected ECM combinations under the TMY and compare its rankings with the results of rankings in future years for the four buildings and listed the ranking changes in Table 11. The ranking changes in the table are calculated by the following function:

$$RCECM_P = \frac{\sum_{i=1}^{n} (R_{i,future} - R_{i,TMY})}{n} \quad (20)$$

where $RCECM_P$ is the rank change of ECM with certain parameter value P; n is the total number of ECM combinations that have the ECM parameter P; $R_{i,future}$ is the ranking of the ith ECM combinations that has parameter P by total BEU in the future; $R_{i,TMY}$ is the ranking of the ith ECM combinations that has parameter P by total BEU in single TMY

year.

For office buildings, natural ventilation is a good measure to save energy and produce financial return and is more welcomed in future climate. The upgrade of the high-efficiency lightings has better rankings compared with TMY climate condition. For residential buildings, the improvement of heating efficiency is more needed for both cities. Higher roof insulations is more needed for residential buildings in both cities in the future, but not recommended for the office building in San Francisco.

For buildings in San Francisco, the most important finding is the change in air infiltration rate rankings. Given that the air infiltration rate is the most important factor influencing the BEU as indicated by feature importance analysis, the magnitude in its ranking changes could alter the picture of future retrofit decision-making with regard to NPV. As shown in Table 11, both residential and office building in San Francisco tend to be more preferring the air infiltration rate of $0.7\,h^{-1}$. For the office building in Philadelphia, a lower air infiltration rate ($0.3\,h^{-1}$) is preferred in the future to ensure the air tightness when the buildings do not need much air exchange. Air exchange is only needed when it is suitable for a natural ventilation to happen (e.g. when cooling is needed and the outdoor air quality meets certain requirement).

Moreover, windows with low SHGC is more preferred for all the four buildings in the future compared to TMY condition. It is needed to gain less heat from the sun because the change in climate in future may incur more cooling need. In the meantime, a lower U-value of the window is also preferable to better insulate the building to reduce the heating energy use for residential buildings and produce more economic profit. The decrease in the rankings of high U-value and high SHGC window type indicates that the future climate does not encourage the thermal exchange between the buildings and the outdoor environment economic-wisely.

It should be given notice that wall insulation is not among the top ECMs for all the four types of buildings in this research. One of the reasons for this might be the pricing of the wall insulation materials, but a more insightful reasoning would be that the wall insulation is bringing both good and bad to the buildings during winter and summer time. A more insulated building would lose less heat during cold winter while also build up heat during hot summer, and the energy saving benefits in dollars does not exceed the comparatively high initial investment and maintenance fee during the lifecycle.

There are also many discrepancies in the strategies for the building to be more economically and environmentally sustainable in the future climate compared with TMY. For example, window shading is less recommended to be applied to office building in San Francisco while it is more recommended to be installed externally on office building in Philadelphia. There is no regular pattern for the choosing of the optimal cooling supply air temperature unless detailed simulation and analysis are conducted for each building. In conclusion, the change of preference for ECM parameters in building retrofit in terms of a high lifecycle NPV differs from building to building, place to place, and it needs careful and insightful analysis and reasoning for different building cases. Therefore, a deep building energy retrofit project should include the factor of GCC impacts on a specific target building using the method and framework suggested in this research.

5. Conclusions

In this research, a framework and method are proposed to evaluate the impacts of different retrofit options to reference buildings in the future climate. A Python retrofit tool is developed to perform parametric study by running EnergyPlus for different retrofit options. Faced with the challenges described in Section 1.2, we developed an approach to reduce the complexity of the training database before the training database is fully generated by means of using LHS method and a feature selection method (JMIM), which finds the most influential ECMs to the NPV of the retrofit during its lifecycle. Highest NPV from the ECMs for the target building is identified and the computational cost in model training and prediction is greatly decreased.

We have validated the feature-selection method by testing it on two referential residential and office building in two different climate zones. RF model is trained by the EnergyPlus hourly time step results simulated by future extreme year weather data to generate the data-driven model, which predicts the future years' hourly BEU during the post retrofit stage. The fidelity of the data-driven method is validated by randomly sampling ECM combinations and comparing future years' BEU from the prediction of the reference model (EnergyPlus). Finally, the ranking of the lifecycle NPV in the future years is compared with that under the current weather condition, and the impacts of choosing different ECM combinations and GCC on building's retrofit strategy in terms of NPV is assessed.

The major findings of this study by implementing the proposed method on four test case buildings located in Philadelphia and San Francisco are:

- Using about 4% samples from the whole combinatorial space suffices to find the approximate most influential ECMs to existing buildings in terms of specific objective function.
- The performance of the trained RF model using the constructed future extreme year's weather in predicting future year's hourly utility is validated.
- GCC does influence the performance of ECMs on BEU and NPV in the future and their rankings in the combinatorial space, and its influence varies from building to building, climate to climate.
- The optimal retrofit strategy of selecting the best ECM combinations under current climate condition is subject to potential change in future climate condition.

In this paper, the proposed method provides a solution to circumvent the huge computational resource needed for BES under future years' weather conditions by using white models like EnergyPlus. Limitation still remains in this method: this paper is considering ten ECM options and each with limited number of parameter options. The results shown in this paper indicate that for a building retrofit project, given a number of ECM options, the proposed method is able to purvey a reasonable guide for screening out the least influential ECMs and finding the design strategy towards optimal with a given objective. Since NPV is the most important factor in the decision making in building retrofit studies, this research is a first step that provides a method to tackle the obstacle of huge computational costs and illustrates the importance of conducting building retrofit analysis under climate change and proves that the climate change has distinct impact on the BEU and NPV when different ECM combinations are applied to a building in the retrofit analysis compared with current climate. It is pointed out that the impact of climate change on the optimization procedure of future retrofit projects should be paid attention as it adds great complexity to the decision-making models for deep building energy retrofit.

References

[1] Srinivasan Ravi S, Braham William W, Campbell Daniel E, Curcija CD. Re(De)fining Net Zero Energy: Renewable Emergy Balance in environmental building design. Build Environ 2012;47:300–15.
[2] Asadi E, da Silva MG, Antunes CH, Dias L. Multi-objective optimization for building retrofit strategies: A model and an application. Energy Build 2012;44:81–7.
[3] Asadi E, da Silva MG, Antunes CH, Dias L. A multi-objective optimization model for building retrofit strategies using TRNSYS simulations, GenOpt and MATLAB. Build Environ 2012;56:370–8.
[4] Asadi E, Silva MGd, Antunes CH, Dias L, Glicksman L. Multi-objective optimization for building retrofit: A model using genetic algorithm and artificial neural network and an application. Energy Build 2014;81:444–56.
[5] Chuah JW, Raghunathan A, Jha NK. ROBESim: A retrofit-oriented building energy simulator based on EnergyPlus. Energy Build 2013;66:88–103.
[6] Lee SH, Hong T, Piette MA, Sawaya G, Chen Y, Taylor-Lange SC. Accelerating the

energy retrofit of commercial buildings using a database of energy efficiency performance. Energy 2015;90, Part 1:738–47.
[7] Levine MD, Feng Wei, Jing Ke, Hong Tianzhen, Zhou N. A retrofit tool for improving energy efficiency of commercial buildings. In: Levine MD, editor. Berkeley, CA: Lawrence Berkeley National Laboratory; 2014.
[8] Rysanek AM, Choudhary R. Optimum building energy retrofits under technical and economic uncertainty. Energy Build 2013;57:324–37.
[9] Yalcintas M. Energy-savings predictions for building-equipment retrofits. Energy Build 2008;40:2111–20.
[10] Hong T, Piette MA, Chen Y, Lee SH, Taylor-Lange SC, Zhang R, et al. Commercial Building Energy Saver: An energy retrofit analysis toolkit. Appl Energy 2015;159:298–309.
[11] Chen Y, Hong T, Piette MA. Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. Appl Energy 2017;205:323–35.
[12] Chen Y, Hong T. Impacts of building geometry modeling methods on the simulation results of urban building energy models. Appl Energy 2018;215:717–35.
[13] IPCC. Contribution of working group III to the fourth assessment report of the intergovernmental panel on climate change, 2007. United Kingdom and New York, NY, USA: Cambridge University Press, Cambridge; 2007.
[14] Wang X, Chen D, Ren Z. Global warming and its implication to emission reduction strategies for residential buildings. Build Environ 2011;46(4):871–83.
[15] Radhi H. Evaluating the potential impact of global warming on the UAE residential buildings – A contribution to reduce the CO_2 emissions. Build Environ 2009;44(12):2451–62.
[16] Shen P. Impacts of climate change on U.S. building energy use by using downscaled hourly future weather data. Energy Build 2017;134:61–70.
[17] Shen P, Lukes JR. Impact of global warming on performance of ground source heat pumps in US climate zones. Energy Convers Manage 2015;101:632–43.
[18] Shen P, Lior N. Vulnerability to climate change impacts of present renewable energy systems designed for achieving net-zero energy buildings. Energy 2016;114:1288–305.
[19] Sun Y, Huang P, Huang G. A multi-criteria system design optimization for net zero energy buildings under uncertainties. Energy Build 2015;97:196–204.
[20] Wang B, Xia X, Zhang J. A multi-objective optimization model for the life-cycle cost analysis and retrofitting planning of buildings. Energy Build 2014;77:227–35.
[21] Lu Y, Wang S, Zhao Y, Yan C. Renewable energy system optimization of low/zero energy buildings using single-objective and multi-objective optimization methods. Energy Build 2015;89:61–75.
[22] LBNL, EnergyPlus (version 8.5.0) [Software]. In: Lawrence Berkeley National Laboratory, http://www.eere.energy.gov/buildings/energyplus/, 2015.
[23] Hirsch JJ. eQuest. In: L.B.N. Laboratory (Ed.), U.S.; 2016.
[24] Klein SA, et al. TRNSYS 17: A transient system simulation program. Madison, USA: Solar Energy Laboratory, University of Wisconsin; 2010.
[25] Wetter M, Wright J. A comparison of deterministic and probabilistic optimization algorithms for nonsmooth simulation-based optimization. Build Environ 2004;39(8):989–99.
[26] Michael W, Jonathan AW. Comparison of a generalized pattern search and a genetic algorithm optimization method. Proc. of the 8th IBPSA conference, Eindhoven, Netherlands. 2003.
[27] Chidiac SE, Catania EJC, Morofsky E, Foo S. Effectiveness of single and multiple energy retrofit measures on the energy consumption of office buildings. Energy 2011;36(8):5037–52.
[28] Eisenhower B, O'Neill Z, Narayanan S, Fonoberov VA, Mezić I. A methodology for meta-model based optimization in building energy models. Energy Build 2012;47:292–301.
[29] Chow DHC, Li Z, Darkwa J. The effectiveness of retrofitting existing public buildings in face of future climate change in the hot summer cold winter region of China. Energy Build 2013;57:176–86.
[30] Nik VM, Mata É, Sasic Kalagasidis A. A statistical method for assessing retrofitting measures of buildings and ranking their robustness against climate change. Energy Build 2015;88:262–75.
[31] DOE. The Encyclopedic Reference to EnergyPlus Input and Output. EnergyPlus Development Team; 2014.
[32] Shao Y, Geyer P, Lang W. Integrating requirement analysis and multi-objective optimization for office building energy retrofit strategies. Energy Build 2014;82:356–68.
[33] N.R.E. Laboratory. National residential efficiency measures database. Golden, CO: N.R.E. Laboratory; 2013.
[34] DOE. Building component cost community. DOE; 2015.
[35] Pope VD, Gallani ML, Rowntree PR, Stratton RA. The impact of new physical parameterizations in the Hadley Centre climate model - HadAM3. Clim Dyn 2000;16(2–3):123–46.
[36] Gaterell MR, McEvoy ME. The impact of climate change uncertainties on the performance of energy efficiency measures applied to dwellings. Energy Build 2005;37(9):982–95.
[37] Chan ALS. Developing future hourly weather files for studying the impact of climate change on building energy performance in Hong Kong. Energy Build 2011;43(10):2860–8.
[38] IPCC. Climate change 2013– the physical science basis, fifth assessment report of the IPCC. In: Stocker T.F, Qin D, editors. New York; 2013.
[39] Belcher SE, Hacker JN, Powell DS. Constructing design weather data for future climates. Building Serv Eng Res Technol 2005;26(1):49–61.
[40] NOAA. NowData – online weather data. National Oceanic and Atmospheric Administration; 2016.
[41] Bennasar M, Hicks Y, Setchi R. Feature selection using Joint Mutual Information Maximisation. Expert Syst Appl 2015;42(22):8520–32.
[42] Shannon CE. A Mathematical Theory of Communication. Bell Syst Tech J 1948;27:379–423.
[43] Mitchell TM. The need for biases in learning generalizations, CBM-TR 5-110. New Brunswick, New Jersey, USA: Rutgers University; 1980.
[44] Quinlan JR. Induction of decision trees. Mach Learn 1986;1(1):81–106.